Express) Mail No. EV417055083US

PCT

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification [4]: C12N 15/00, C07H 15/12 C12N 1/20	A1	(11) International Publication Number: **WO 89/ 00604**
		(43) International Publication Date: 26 January 1989 (26.01.89)

(54) Title: METHOD FOR IMPROVING TRANSLATION EFFICIENCY
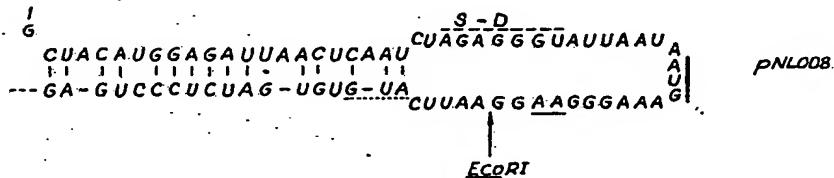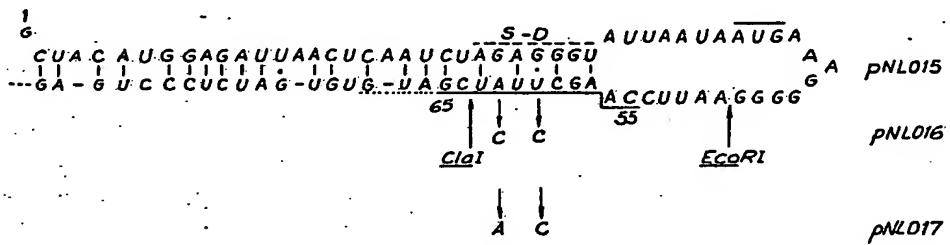
(57) Abstract

A method for increasing the translation efficiency of a mRNA sequence is provided. In certain of its preferred embodiments, the method comprises the steps of: (a) constructing a predicted secondary structure for the mRNA; (b) analyzing the predicted secondary structure to determine if either or both of the AUG initiation codon and the Shine-Dalgarno sequence is contained in a double stranded portion of a stem-loop region of the predicted secondary structure; (c) calculating a free energy value for the stem-loop region; and (d) if the calculated free energy value is in the range of from zero to about -10.0 kcal/mole, modifying the DNA sequence for the mRNA so that when the modified sequence is transcribed it produces a modified mRNA sequence which has a predicted secondary structure wherein the AUG initiation codon and the Shine-Dalgarno sequence are not included in a double stranded portion of a stem-loop region of the predicted secondary structure. By means of this method, ten-fold increases in protein production have been achieved.

METHOD FOR IMPROVING TRANSLATION EFFICIENCY

BACKGROUND OF THE INVENTION

1. Field of the Invention

This invention relates to a method for increasing the production of proteins by biological cells and, in particular, to a method for improving the efficiency with which messenger RNA (mRNA) is translated.

2. Description of the Prior Art

Extensive research has been performed to attempt to understand why some messenger RNAs are translated more efficiently than other messenger RNAs. For example, studies using E. coli have been performed in which nucleotide sequences of varying lengths have been placed between the AUG initiation codon and the Shine-Dalgarno (SD) sequence. See Shepard et al., "Increased Synthesis in E. coli of Fibroblast and Leukocyte Interferons Through Alterations in Ribosome Binding Sites," DNA, 1:125-131 (1982). These studies identified an optimal spacing of 9 nucleotides. (As known in the art, the Shine-Dalgarno sequence comprises 3-12 nucleotides which are found in prokaryotic mRNAs and which are complementary to the 3' end of 16s rRNA.)

Similarly, the effects of base composition in the region between the SD sequence and the initiation codon have been studied. See De Boer et al., "Portable Shine-Dalgarno Regions: A System for a Systematic Study of Defined Alterations of Nucleotide Sequences

-2-

within E. coli Ribosome Binding Sites," DNA, 2:231-235 (1983). In this case, it was found that the presence of A or T residues resulted in higher levels of mRNA translation, while C or G residues led to lower levels of translation.

In addition to the foregoing, various studies have been performed to determine the effects of mRNA secondary structure on translation efficiency. See, for example, Boyen et al., "Enhancement of Translation Efficiency in Escherichia coli by Mutations in a Proximal Domain of Messenger RNA," J. Mol. Biol., 162:715-720 (1982); Coleman et al., "Mutations Upstream of the Ribosome-binding Site Affect Translational Efficiency," J. Mol. Biol., 181:139-143 (1985); Hall et al., "A Role for mRNA Secondary Structure in the Control of Translation Initiation," Nature, 295:616-618 (1982); Iserentant et al., "Secondary Structure of mRNA and Efficiency of Translation Initiation," Gene, 9:1-12 (1980); Kastelein et al., "Effect of the sequences upstream from the ribosome-binding site on the yield of protein from the cloned gene for phage MS2 coat protein," Gene, 23:245-254 (1983); Queen et al., "Differential Translation Efficiency Explains Discoordinate Expression of the Galactose Operon," Cell, 25:241-249 (1981); Schoner et al., "Role of mRNA translational efficiency in bovine growth hormone expression in E. coli," Proc. Nat. Acad. Sci., USA, 81:5403-5407 (1984); Shepard et al., supra; and Steege,

D., "5'-Terminal nucleotide sequence of Escherichia coli lactose repressor mRNA: Features of translational initiation and reinitiation sites," Proc. Natl. Acad. Sci., USA, 74:4163-4167 (1977).

In these studies, A-U and G-C base pairing has been used to construct mRNA secondary structures from mRNA primary structures. The secondary structures have then been examined to determine if either or both of the translation initiation determinants, e.g., the SD sequence and the initiation codon, appear in a double stranded portion of the folded mRNA molecule, as opposed to a single stranded portion. In addition, the thermodynamic stabilities of the secondary structures have been calculated using the methods of Tinoco et al. See Tinoco et al., "Improved Estimation of Secondary Structure in Ribonucleic Acids," Nature New Biology, 246:40-41 (1973); Tinoco et al., "Estimation of Secondary Structure in Ribonucleic Acids", Nature, 230:362-367 (1971); and Borer et al., "Stability of Ribonucleic Acid Double-Stranded Helices," J. Mol. Biol., 86:843-853 (1974). See also Salser, W., Cold Spring Harbor Symp. Quant. Biol., 42:985-1002 (1977). Lower levels of protein production have generally been found to correlate with the sequestering of translation initiation determinants in double stranded regions of thermodynamically stable secondary structures, i.e., secondary structures having a calculated free energy (ΔG) more negative than about -10.0 kcal/mole, i.e.,

-4-

free energies more negative than the free energy of
hydrolysis of ATP ($\Delta G = -7.3$ kcal/mole).

Significantly, with regard to the present
invention, the sequestering of translation initiation
determinants in double stranded regions of secondary
structures having calculated free energies more
positive than -10.0 kcal/mole have not been considered
important with regard to changes in translation
efficiency. For example, in Hall et al., supra, a SD
sequence sequestered in a secondary structure having a
free energy of -2.0 kcal was considered accessible for
ribosome binding due to the instability of the
secondary structure. Similarly, in Iserentant et al.,
supra, an SD sequence was considered largely single
stranded where release of the sequence from a double
strand configuration only required 3.2 kcal. Along
these same lines, in De Boer et al., supra, it was
stated that the observed differences in translation
efficiency which these workers had found could not be
explained on the basis of secondary structure because
significantly stable stem-loops having ΔG values less
than -10 kcal could not be found. Again, in Coleman et
al., supra, a secondary structure having a free energy
of -3.05 kcal was considered unlikely to significantly
affect ribosome binding to mRNA.

In contrast to these views of prior workers, as
discussed and demonstrated in detail below, in
accordance with the present invention it has been

-5-

surprisingly found that translation efficiency can be significantly affected, e.g., by a factor of 10, by the sequestering of translation initiation determinants in secondary structures having calculated free energies more positive than -10.0 kcal/mole, e.g., on the order of -3.9 kcal/mole. As shown below, based on this discovery, even greater increases in protein production can be achieved than would have been achieved based on the prior understanding of the effects of mRNA secondary structure on translation efficiency.

SUMMARY OF THE INVENTION

In view of the foregoing state of the art, it is an object of this invention to increase the production of protein by biological cells. More particularly, it is an object of the invention to increase the efficiency with which mRNA is translated to produce protein.

To achieve the foregoing and other objects, the invention provides a method for increasing the translation efficiency of a mRNA sequence which is produced from a DNA or RNA sequence comprising the steps of:

(a) determining and analyzing a predicted secondary structure for the mRNA to determine if either or both of the AUG initiation codon and the Shine-Dalgarno sequence is contained in a double stranded portion of a stem-loop region of the predicted secondary structure;

-6-

(b) calculating a free energy value for the stem-loop region; and

(c) modifying the DNA or RNA sequence for the mRNA, if the calculated free energy value is in the range of from zero to about -10.0 kcal/mole, so that the modified DNA or RNA sequence produces a modified mRNA sequence which has a predicted secondary structure wherein either:

(i) the AUG initiation codon and the Shine-Dalgarno sequence are not included in a double stranded portion of a stem-loop region of the predicted secondary structure; or

(ii) if either or both of the AUG initiation codon and the Shine-Dalgarno sequence are included in a double stranded portion of a stem-loop region of the predicted secondary structure, the calculated free energy value for such stem-loop region is more positive than the free energy value calculated in step (b).

As described in detail below, by means of the foregoing procedure, ten-fold increases in protein production have been achieved.

In certain preferred embodiments, the free energy value calculated in step (b) is in the range of from zero to about -7.0 kcal/mole, i.e., the calculated free energy is more positive than the free energy of hydrolysis of ATP. This range is particularly likely to have been ignored by prior art workers since it

-7-

represents structures whose energy is less than the energy of hydrolysis of only one ATP.

The accompanying drawings, which are incorporated in and constitute part of the specification, illustrate

5   the application of the invention to the production of human α1 interferon, and together with the description, serve to explain the principles of the invention.

10

15

20

25

-8-

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the construction of hybrid plasmid pNL014. Abbreviations used: lpp ^{P} = E. coli lipoprotein promoter, lpp ^{t} = E. coli lipoprotein transcription terminator, A^r = Ampicillin resistance gene, Gal ^{P} = Yeast galactose promoter, α1 = mature interferon α1 gene, Met-α1 = methionyl-α1, α1 ^{t} = Interferon α1 transcription terminator, SUC ^{t} = Yeast invertase transcription terminator, Xb = XbaI, E = EcoRI, H = HindIII, S = SalI, C = ClaI, 2u = Yeast 2u replication origin, URA3 = Yeast URA3 gene, and a.a. = amino acid.

Figure 2 shows the construction of hybrid plasmid pNL015. Abbreviations: as in Figure 1. pIN-I = pIN-I-A$_3$ vector. Marked bases are those which do not appear in pNL008.

Figure 3 shows the construction of hybrid plasmid pNL008. Abbreviations: as in Figure 1. S1 = S1 nuclease. Marked bases are those which do not appear in pNL015. EcoRI linker (GAATTC) was obtained from New England Biolabs.

Figure 4 shows the predicted secondary structures for mRNA produced from plasmids pNL015 and pNL008. The sequences start with the first base (1) of the transcripts. Sequences under the broken lines are the Shine-Dalgarno region. The initiation codon is indicated by a heavy bar and the deleted or inserted sequences are marked by a light bar. The AUG of the

-9-

methionyl interferon is marked by dots. Calculated free energies of the secondary structures are -3.9 kcal/mole and - 3.2 kcal/mole for pNL015 and pNL008 transcripts, respectively. Arrows indicate base substitutions as occurred in pNL016 and pNL017.

Figure 5 shows the results of a pulse-chase analysis of the in vivo stability of interferon fusion protein produced by JA221/pNL008 (lanes 1 - 4) and JA221/pNL015 (lanes 5-8). After E. coli cells were incubated with [^{35}S]- methionine for 40 minutes, a 40,000-fold excess of unlabeled methionine was added. At various chase times, samples were withdrawn and cell lysates were prepared for PAGE and immunostaining. Lane 0 contains molecular weight standards. Chase times were -- lanes 1 and 5: 0 min.; lanes 2 and 6: 15 min.; lanes 3 and 7: 30 min.; lanes 4 and 8: 60 min. Alpha 1 indicates the position of IFN αl fusion protein. The M.W. of IFN fusion protein from pNL015 was slightly higher than that from pNL008 due to 3 extra amino acid residues.

Figure 6 shows the results of RNA dot blot hybridization experiments for JA221/pNL015 and JA221/pNL008. Columns 1 and 2 represent serial dilution (1:5) of total cellular RNA (top row 40 ug) of JA221/pNL015 and JA221/pNL008, respectively. The total RNA was isolated following the procedure of Young and Furano and applied to nitrocellulose paper. See Young, F. S. and Furano, A. V., "Regulation of the synthesis

-10-

of E. coli elongation factor Tu," Cell, 24 (1981) 695-706. Nick-translated EcoRI fragment of IFN αl gene from pNL015 was used for hybridization according to Maniatis et.al.. See Maniatis, T., Fritsch, E. F., and Sambrook, J., Molecular Cloning, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, (1982).

Figure 7 shows the results of stability analyses of IFN αl mRNA transcribed from pNL015 (solid triangles) and pNL008 (open triangles). Rifampicin (Calbiochem) was added (200ug/ml) to five 15 ml cell cultures of JA221/pNL008 and JA221/pNL015, respectively. The cultures were incubated at 37°C for 5, 10, 15, and 20 min. After each incubation, 50 uCi of [^{35}S]-methionine was added to each culture, and the mixtures were incubated for another 1.5 min. The control cultures (zero time) were labeled with [^{35}S]-methionine in the absence of rifampicin. Cell extracts were prepared and analyzed by SDS-PAGE. The IFN αl peak positions were identified by immunostaining. The peaks were cut out from the gel blot and the radioactivities of these fractions were measured in ACS scintillation fluid after being solubilized in 0.5 ml of a NCS solubilizer (Amersham Corp.) - water mixture (9:1) at 50°C for 2 hrs. The incorporation rates were calculated as percentage of incorporation at zero time.

Figure 8 shows the construction of pNL016 and pNL017. Abbreviations as in Figure I. Oligonucleotides were synthesized by using Applied Biosystems DNA

-11-

Synthesizer and purified by gel elution. Following phosphorylation by polynucleotide kinase, appropriate pairs of oligonucleotides were mixed and annealed at 70°C in the presence of 66 mM Tris.HCl, pH 7.5 / 6 mM MgCl$_2$ / 500 mM ATP in a volume of 270 ul. The mixtures were cooled to 15°C during one hour to form DNA fragment A or B. Arrows indicate positions of base substitution as compared to pNL015.

5

10

15

20

25

-12-

DESCRIPTION OF THE PREFERRED EMBODIMENTS

As described above, the present invention relates to the discovery that the sequestering of one or both of the AUG initiation codon and the Shine-Dalgarno sequence in a double stranded portion of even a weakly bound (relatively unstable) mRNA secondary structure, i.e., a secondary structure having a calculated free energy more positive than -10.0 kcal/mole, can have significant effects on protein production. In particular, it has been found that the elimination of such secondary structures can result in over a five-fold increase, e.g., a ten-fold increase, in protein production.

The first step of the process of the invention involves determining if all or part of either or both of the AUG initiation codon and the Shine-Dalgarno sequence is contained in a double stranded portion of a stem-loop region of the predicted secondary structure for the mRNA (i.e., the secondary structure predicted from pairing of the bases of the mRNA's primary structure as opposed to an experimentally observed secondary structure).

This determination is most conveniently performed by conducting computer analyses on the base sequence of the mRNA to identify one or, in some cases, a group of possible secondary structures for the mRNA. When a group of structures is identified, a most probable secondary structure can normally be selected following

techniques known in the art. See Akiyoshi Wada and Akira Suyama, "Local Stability of DNA and RNA Secondary Structure and its Relation to Biological Functions" Prog. Biophys. Molec. Biol., Vol. 47, pp 113-157

5 (1986); Nicholls, N., The Regulation of Bacillus Licheniformis Penicillinase: Locating and Sequencing the Repressor Gene, Ph.D. Thesis, Rutgers University, Dissertation Abstracts International, Molecular Biology, Vol. 47/06-B, 1986. In some cases, however,

10 it may be advantageous to further analyze a group of possible secondary structures, as opposed to a single most probable structure. For example, depending on the details of the sequences involved, it may be possible to produce a modified mRNA which eliminates binding of

15 the AUG initiation codon and/or the Shine-Dalgarno sequence for a whole family of secondary structures. As an alternate to using a computer program, secondary structures can be determined visually, i.e., by examining the primary structure and manually aligning

20 A-U and C-G pairs, and the free energies of such structures can be manually calculated using, for example, the Tinoco techniques, supra.

Examples of parameters which can be considered in selecting most probable secondary structures include

25 percent base match, probability of finding as good a match in a random sequence of bases of the same length, and secondary structure free energy. With regard to free energy, it is important not to dismiss secondary

-14-

structures having calculated free energies more positive than -10.0 kcal/mole as prior workers have done (see above), since as shown below, such weak secondary structures can significantly affect protein production.

A suitable computer program for determining mRNA secondary structures is the SEQ - DNA Sequence Analysis System marketed by IntelliGenetics, Inc., of Palo Alto, California (hereinafter the "SEQ program"). A description of this program can be found in Brutlag, et al., "SEQ: A Nucleotide Sequence Analysis and Recombination System," Nucleic Acids Research, 10:279-294 (1982). As discussed therein, the SEQ program uses the Tinoco et al. techniques, supra, to calculate free energies. See also Zuker and Sankov, Bull. of Math. Biol., 46:591-621 (1984).

Once the predicted secondary structure or structures have been constructed, they are analyzed to determine: 1) if all or part of either or both of the AUG initiation codon and the Shine-Dalgarno sequence is contained in a double stranded portion of a stem-loop region of the secondary structure or structures; and 2) if the free energy of such a stem-loop region is between zero and about -10.0 kcal/mole. The free energy can be conveniently calculated using a computer program, such as, the SEQ program.

If both of the foregoing conditions are satisfied, the mRNA sequence is then analyzed to identify

modifications to the mRNA sequence, i.e., base additions, deletions, and/or substitutions, which will result in predicted secondary structures wherein the AUG initiation codon and the Shine-Dalgarno sequence are not contained in a double stranded portion of the predicted secondary structure. Alternatively, but less preferably, the sequence is analyzed for modifications which will place the AUG initiation codon and/or the Shine-Dalgarno sequence in double stranded portion of a predicted secondary structure which is even less stable than the predicted secondary structure of the original mRNA sequence, i.e., in a secondary structure whose calculated free energy is more positive than the calculated free energy of the original mRNA sequence.

Various constraints must be kept in mind in considering possible modifications to the mRNA sequence. For example, if the protein which is to be produced is to remain unchanged, modifications which are to be made to the portion of the mRNA sequence which codes for the protein are limited to those which degenerately code for the same amino acids. See, for example, Nussinov, R., "RNA Folding Is Unaffected by the Nonrandom Degenerate Codon Choice", Biochimica et Biophysica Acta, 698:111-115 (1982). Also, when modifying the protein coding region, it is preferred to select codons which correspond to those tRNAs which are most abundant in the host cell which will be used for protein production.

-16-

When modifying non-coding regions of the mRNA sequence, other constraints come into play. For example, it is in general preferred to avoid changes in the Shine-Dalgarno sequence and the spacing between the AUG initiation codon and the Shine-Dalgarno sequence that may adversely affect the translation initiation process. Also, as reported by De Boer et al., the addition of C or G residues to the region between the initiation codon and the Shine-Dalgarno sequence is not preferred. Of course, although changes of the foregoing types are not preferred, they can be used if necessary to achieve the desired predicted secondary structure.

Once a modified mRNA sequence has been selected, its predicted secondary structure is determined following the same procedures as those used for the original mRNA sequence. Again, the structure, or in some cases, structures are examined to determine the locations of the AUG initiation codon and the Shine-Dalgarno sequence, and, if necessary, the free energy of the region of the secondary structure which contains these elements is calculated. If necessary, different or further modifications of the original mRNA sequence are then analyzed until a modified mRNA sequence is selected which achieves the goal of minimizing the likelihood that secondary structure will interfere with the functioning of the Shine-Dalgarno sequence and the AUG initiation codon.

-17-

Production of the modified mRNA sequence is achieved by altering the DNA or RNA sequence which codes for the original mRNA. Various techniques can be used to produce the modified DNA or RNA sequence. For example, site-specific mutagenesis can be used to achieve the modifications. See, for example, Messing, J., "New M13 Vectors for Cloning," Methods in Enzymology -- Recombinant DNA, Wu et al. (eds.), Academic Press, New York, Volume 101, Part C, pages 20-79 (1983). Similarly, the modifications can be produced through the use of synthetic gene fragments and nucleases which fragment the DNA or RNA sequence at predetermined locations. As will be evident to persons of ordinary skill in the art, other techniques which may be developed in the future for preparing or altering DNA or RNA sequences can be used in the practice of the present invention.

Production of protein using the modified DNA or RNA sequence is achieved following standard recombinant DNA/RNA techniques. A discussion of these techniques can be found in, for example, Molecular Cloning: A Laboratory Manual, by T. Maniatis et al., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (1982). Again, as with the altering of DNA or RNA sequences, the present invention can be used in combination with techniques developed in the future for producing proteins from genetically altered cells.

Without intending to limit it in any manner, the

-18-

invention will be more fully described by the following examples. The materials and methods which are common to the examples are as follows.

MATERIALS AND METHODS

Strains, plasmids, and plasmid isolation

E. coli K-12 strain JA221 (hsdM+ hsdR- recA Leu LacY Trp) was used as the transformation host. pIN series vectors, in particular, pIN-I-A$_2$ and pIN-I-A$_3$, were used as the cloning vehicles. See Nakamura, K., and Inouye, M., "Construction of versatile expression cloning vehicles using the lipoprotein gene of Escherichia coli," The EMBO Journal, 1:771-775 (1982). These vectors use the promoter and the 5' untranslated region of the E. coli outer membrane lipoprotein gene for transcriptional and translational initiation of the cloned gene.

Plasmid pCGS282, which was obtained from Collaborative Research, Inc., Lexington, Massachusetts, was used as a source of a leukocyte interferon α1 gene (see Figure 1). This plasmid is a hybrid plasmid in which a mature human interferon α1 gene has been inserted between the S. cerevisae galactose promoter (GAL^P) and the S. cerevisae invertase transcription terminator (SUC^t). In this construction, 69 base pairs coding for the signal sequence of the prointerferon protein have been removed and replaced by an ATG translation initiation codon and Cla I and Hind III linkers.

-19-

Plasmids were isolated by the method described by Tanaka and Weisblum. Tanaka, T. and Weisblum, B., "Construction of a colicin El-R factor composite plasmid in vitro: Means of amplification of deoxyribonucleic acid," J. Bacteriol., 121:345-362 (1975).

Preparation of cell extracts, gel electrophoresis and immunostaining.

Cell pellets from 20 ml of log phase E. coli cell culture were suspended in 2 ml of 0.1 M phosphate buffer (pH 7.5). Cells were broken by sonication for 5 min. Cell extracts were obtained after 1 hour ultracentrifugation at 40,000 rpm at 4°C.

Sodium dodecylsulfate-polyacrylamide gel electrophoresis (SDS - PAGE) was performed using a 15% slab gel by the method of Anderson et al. Anderson, C. W., Baum, P. R. and Gesteland, R.F., "Processing of adenovirus 2-induced proteins," J. Virol., 12 (1973) 241-252.

A mixture of 20 ul of cell extract and 20 ul of loading buffer were loaded on the 15% SDS - PAGE and the samples were electrophoresed at 140 V until the bromophenol blue dye moved out of the gel. The protein bands on the gel were transferred to a nitrocellulose sheet by electroblotting at 70 V for 3 hours at 4°C. The sheet was washed with TBS (20 mM Tris-HCl, pH 7.5; 500 mM NaCl) and blocked with TBS containing 3% BSA at room temperature for 30 min. The blocked sheet was

-20-

incubated overnight in T-TBS (TBS + 0.05% Tween 20) at
4°C containing a 100-fold dilution of rabbit polyclonal
antibody against alpha interferon which was obtained
from Interferon Sciences, Inc., New Brunswick, New
5   Jersey. The sheet was washed three times with T-TBS,
incubated in a 3000-fold dilution of BioRad peroxidase
conjugated goat-anti-rabbit IgG at room temperature for
2 hours, and washed three times with TBS. The protein
bands were visualized by developing the sheet in a
10   freshly prepared solution of 0.05 % 4-chloro-1-
naphthol/0.00015 % H_2O_2 at room temperature for 30 min.
The developed sheet was washed four to five times with
distilled water to stop the reaction and then air
dried.

15   Example 1

Construction of Plasmid pNL015

This example relates to the construction of a
plasmid (pNL015) which produces a mRNA sequence in
which the Shine-Dalgarno sequence is contained in a
20   double stranded portion of a stem-loop region of the
sequence's predicted secondary structure. The
calculated free-energy of the stem-loop region is -3.9
kcal/mole, i.e., the calculated free-energy is in the
range of free energies which prior workers in the art
25   thought could not significantly affect protein
production.

pNL015 was constructed by first constructing
pNL014 by ligating a HindIII - SalI DNA fragment

containing the IFN αl gene from pCGS282 to the large HindIII − SalI fragment of pIN-I-A$_3$ (see Figure 1). In this construction, the promoter, the 5' untranslated region of the lipoprotein gene, a sequence coding for the first two amino acid residues of the prolipoprotein, and a linker sequence coding for seven amino acid residues are situated 5' to the coding region of the methionyl IFN αl gene. At the 3' end, the IFN αl 3' untranslated region is followed by an invertase transcription terminator. Accordingly, there are two transcriptional termination sequences, both eukaryotic in nature, following the αl interferon coding sequence.

The biological activity of interferon isolated from E. coli cells JA221 harboring pNL014 was measured using Vesicular Stomatitis virus (Indiana Strain) on HEp-2 cells in a cytopathic effect assay. See Lee, N., Cozzitorto, J., Wainwright, N. and Testa, D., "Cloning with tandem gene systems for high level gene expression," Nucleic Acids Res., 12 (1984) 6797-6812. The quantity of IFN isolated from these cells was in the range of 1.2 x 10^4 units/ml/OD.

pNL015 was constructed from pNL014 and pIN-I-A$_2$ following the procedures shown in Figure 2. In this plasmid, the two eukaryotic terminators of plasmid pNL014 are replaced by a single E. coli lipoprotein terminator. This change was found to result in a two

-22-

fold increase in interferon activity (i.e., to the range of 2.8 x 10^4 units/ml/OD).

Example 2

Construction of a Predicted Secondary

5   Structure for mRNA Produced from pNL015

A predicted secondary structure for mRNA produced from pNL015 was constructed using the IntelliGenetics SEQ program, supra. The parameter settings employed in performing the computer analysis of the primary mRNA

10   sequence were the default parameters for the SEQ program, i.e., Percentmatch = 70%, AfterMismatch = 2, LoopOut = 3, MinLoop = 3, and MaxLoop =50. Various combinations of the Percentmatch, MaxLoop, MinLoop, and AfterMismatch parameters used in the SEQ program were

15   found to predict the same secondary structure as the default parameters.

The predicted secondary structure in the region of the AUG initiation codon and the Shine-Dalgarno sequence (AGAGGGU) obtained by this analysis is shown

20   in Figure 4. The secondary structure shown has a free energy of -3.9 kcal/mole, a percentage match of 75%, and an "E" factor, i.e., a probability of finding as good a match in a random sequence of bases of the same length, of 2.387.

25   As is evident from Figure 4, the Shine-Dalgarno sequence is contained in a double stranded portion of a stem-loop region of the predicted secondary structure. Moreover, as indicated above, the stem-loop region has

-23-

a calculated free energy in the range of 0 to -10.0 kcal/mole, i.e., a calculated free energy of -3.9 kcal/mole.

Example 3

Construction of Plasmid pNL008

This examples illustrates the modification of pNL015 to produce a plasmid (pNL008) which produces a mRNA having a predicted secondary structure in which neither the AUG initiation codon nor the Shine-Dalgarno sequence are included in a double stranded portion of a stem-loop region of the secondary structure.

To construct pNL008, pNL015 was linearized with ClaI, blunt-ended with S1 nuclease, ligated with EcoRI linkers and digested with EcoRI. The small EcoRI fragment was isolated and inserted into the EcoRI site of the pIN-I-A$_2$ cloning vehicle as shown in Figure 3.

pNL008 differs from pNL015 in that the mRNA transcript produced by pNL008 has 11 bases deleted starting from base 17 downstream of the AUG initiation codon, and a 2 base (A-A) insertion between bases 9 and 10. Other than these differences, the two plasmids are identical.

Figure 4 shows a predicted secondary structure for pNL008 where the deleted and inserted sequences in pNL015 and pNL008 are indicated by the light bars. The structure for pNL008 shown in Figure 4 was constructed manually from the primary structure for the plasmid's first 74 bases. Using the Tinoco technique, the free

-24-

energy for this structure was calculated manually and found to be -3.2 kcal/mole. Computer analysis using the SEQ program of the same primary sequence predicted a less stable structure having an energy of -1.7 kcal/mole in which the SD sequence was partially contained in a double stranded portion of a stem-loop region.

As is evident from Figure 4, the modifications to pNL015 have freed both the initiation codon and the Shine-Dalgarno sequence from double stranded regions of the predicted secondary structure. Significantly, this freeing was found to result in a ten-fold increase in the production of $\alpha 1$ interferon. Specifically, E. coli cells transformed with pNL008 were found to produce 3×10^5 units/ml/OD as compared to only 2.8×10^4 units/ml/OD for cells transformed with pNL015.

Example 4

Experiments to Confirm that the Difference in Protein Production Between Plasmids pNL008 and pNL015 Is Due To the Predicted Secondary Structures of Their mRNAs

A series of experiments were performed to confirm that the observed difference in protein production between plasmids pNL008 and pNL015 was due to their predicted secondary structures as opposed to other factors.

First, it was determined that the specific activities of interferon encoded by pNL015 and pNL008

-25-

were the same. In addition, it was determined that cells containing pNL015 had a proportionally lower quantity of protein detectable by Western blot analysis than cells containing pNL008. Accordingly, it was concluded that the difference in N-terminus amino acid residues was not a factor affecting the antiviral activity of the interferon protein.

The in vivo stability of the interferon protein was examined in a pulse-chase experiment. The interferon proteins encoded by pNL015 and pNL008 showed no noticeable degradation within 60 minutes (see Figure 5). It was therefore concluded that the differential rate of degradation by proteolytic enzymes was not a contributing factor to the observed difference in IFN expression.

Experiments were carried out to quantitate the interferon mRNA synthesized. Dot blot hybridization experiments showed that pNL008 produced slightly more interferon mRNA than pNL015 (see Figure 6). This minimal change in the efficiency of transcription, however, was not large enough to account for the observed ten fold increase in expression level.

The stability of the IFN mRNAs was studied by measuring levels of labeled IFN at various time intervals after the addition of rifampicin, an inhibitor of RNA synthesis, to cultures of JA221/pNL015 and JA221/pNL008. As can be seen in

Figure 7, functional half-lives of the transcripts produced by both pNL015 and pNL008 are approximately 4 to 5 minutes.

Based on the foregoing experiments, it was concluded that none of specific activity, rate of transcription, protein degradation or mRNA degradation could account for the observed difference in expression between pNL008 and pNL015.

In addition to the foregoing, two additional plasmids (pNL016 and pNL017) were constructed to further confirm that the source of the difference in expression between plasmids pNL008 and pNL015 was mRNA secondary structure. These plasmids were formed by synthesizing two EcoRI - Cla I DNA fragments (fragments A and B) to replace the corresponding sequence in pNL015 (bases 50 to 63 in Figure 4). DNA fragment B contained a single base substitution whereby base 60 of the pNL015 transcript was changed from U to C. DNA fragment A contained an additional base substitution whereby both base 60 (U) and base 62 (A) were changed to C.

These changes do not change the amino acid sequence defined by plasmid pNL015 since each of CUA, CUC, and UUA code for leucine. Moreover, using the SEQ program, it was found that the predicted secondary structures for the mRNA sequences corresponding to pNL016 and pNL017, like the predicted secondary structure corresponding to pNL015, had the

-27-

Shine-Dalgarno sequence in a double stranded portion of a stem-loop region of the secondary structure.

The calculated free energy values for the mRNA stem-loop regions for the three plasmids, however, were different. Specifically, whereas the stem-loop region for pNL015 had a free energy of -3.9 kcal/mole, for pNL017, the free energy was -10.8 kcal/mole, and for pNL016, it was -17 kcal/mole. Significantly, E. coli cells transformed with either pNL016 or pNL017 produced no detectable interferon activity.

Since all three plasmids code for the same amino acid sequence, post-translational factors such as specific activity and protein stability cannot contribute to the change in interferon titer. Moreover, substitution of one or two bases in the coding region is unlikely to alter the overall rate of transcription. Having ruled out these factors, the only possibility left is translational inhibition. Translational efficiency is unlikely to be affected by codon usage, since only one synonymous codon substitution is involved. Moreover, in E. coli, the codon CUC used in pNL016 is a more frequently used codon and the codon CUA used in pNL017 is a less frequently used codon in comparison to the UUA codon used in pNL015. See Maruyama et al., "Codon usage tabulated from the GenBank genetic sequence data," Nucl. Acids Res., Vol. 14 sup., pages R151-R197 (1986).

-28-

Yet, both synonymous substitutions resulted in complete
inhibition of expression.

These additional results for plasmids pNL016 and
pNL017 add further support to the conclusion that the

5   efficiency with which pNL015's mRNA is translated is a
function of its secondary structure notwithstanding the
relative instability of that secondary structure.

10

15

20

25

-29-

What is claimed is:

1. A method for increasing the translation efficiency of a mRNA sequence which (i) codes for a protein, (ii) is produced from a RNA or DNA sequence, and (iii) includes an AUG initiation codon and a Shine-Dalgarno sequence, comprising the steps of:

(a) determining and analyzing a predicted secondary structure for the mRNA sequence to determine if one or both of the AUG initiation codon and the Shine-Dalgarno sequence is contained in a double stranded portion of a stem-loop region of the predicted structure;

(b) calculating a free energy value for the stem-loop region;

(c) if the calculated free energy value is in the range of from zero to about -10.0 kcal/mole:

(i) selecting a modified mRNA sequence;

(ii) determining and analyzing a predicted secondary structure for the modified mRNA sequence to determine if either or both of the AUG initiation codon and the Shine-Dalgarno sequence is contained in a double stranded portion of a stem-loop region of the predicted structure;

(iii) repeating steps (c)(i) and (c)(ii), if necessary, until a modified mRNA sequence is identified in which neither the AUG initiation codon nor the Shine-Dalgarno sequence is

contained in a double stranded portion of a
stem-loop region of the predicted secondary
structure; and

(d) modifying the DNA or RNA sequence so that it
produces the modified mRNA sequence.

2.   The method of Claim 1 wherein the free energy
calculated in step (b) is in the range of from zero to
about -7.0 kcal/mole.

3.   The method of Claim 1 wherein the modification
of the DNA or RNA sequence results in the production of
at least five times more protein than produced by the
unmodified DNA or RNA sequence.

4.   A DNA or RNA sequence which has been modified
by the method of Claim 1.

5.   A transformed host cell which includes the
modified DNA or RNA sequence of Claim 4.

6.   A method for increasing the translation
efficiency of a mRNA sequence which (i) codes for a
protein, (ii) is produced from a DNA or RNA sequence,
and (iii) includes an AUG initiation codon and a
Shine-Dalgarno sequence, comprising the steps of:

(a) determining and analyzing a predicted
secondary structure for the mRNA sequence, to determine
if one or both of the AUG initiation codon and the
Shine-Dalgarno sequence is contained in a double
stranded portion of a stem-loop region of the predicted
structure;

-31-

(b) calculating a free energy value for the stem-loop region;

(c) if the calculated free energy value is in the range of from zero to about -10.0 kcal/mole:

(i) selecting a modified mRNA sequence;

(ii) determining and analyzing a predicted secondary structure for the modified mRNA sequence to determine if either or both of the AUG initiation codon and the Shine-Dalgarno sequence is contained in a double stranded portion of a stem-loop region of the predicted structure;

(iii) if neither the AUG initiation codon nor the Shine-Dalgarno sequence is contained in a double stranded portion of a stem-loop region of the predicted secondary structure for the modified mRNA, modifying the DNA or RNA sequence so that it produces the modified mRNA sequence;

(iv) if either or both of the AUG initiation codon and the Shine-Dalgarno sequence is contained in a double stranded portion of a stem-loop region of the predicted structure for the modified mRNA, calculating a free energy value for the stem-loop region;

(v) if the free energy value calculated in step (c)(iv) is more positive than the free energy value calculated in step (b), modifying the

DNA or RNA sequence so that it produces the modified mRNA sequence;

(vi) if the free energy value calculated in step (c)(iv) is equal to or more negative than the free energy value calculated in step (b), repeating steps (c)(i) and c(ii), and, as appropriate, steps (c)(iii), (c)(iv), and (c)(v), until the condition of (c)(iii) or the condition of (c)(v) is satisfied.

7. The method of Claim 6 wherein the free energy calculated in step (b) is in the range of from zero to about -7.0 kcal/mole.

8. The method of Claim 6 wherein the modification of the DNA or RNA sequence results in the production of at least five times more protein than produced by the unmodified DNA or RNA sequence.

9. A DNA or RNA sequence which has been modified by the method of Claim 6.

10. A transformed host cell which includes the modified DNA or RNA sequence of Claim 9.

11. A method for increasing the translation efficiency of a mRNA sequence which (i) codes for a protein, (ii) is produced from a DNA or RNA sequence, and (iii) includes an AUG initiation codon and a Shine-Dalgarno sequence, comprising the steps of:

(a) determining and analyzing a predicted secondary structure for the mRNA sequence to determine if either or both of the AUG initiation codon and the

Shine-Dalgarno sequence is contained in a double stranded portion of a stem-loop region of the predicted secondary structure;

(b) calculating a free energy value for the stem-loop region; and

(c) modifying the DNA or RNA sequence for the mRNA, if the calculated free energy value is in the range of from zero to about -10.0 kcal/mole, so that when the modified DNA or RNA sequence produces a modified mRNA sequence which has a predicted secondary structure wherein either:

(i) the AUG initiation codon and the Shine-Dalgarno sequence are not included in a double stranded portion of a stem-loop region of the predicted secondary structure; or

(ii) if either or both of the AUG initiation codon and the Shine-Dalgarno sequence are included in a double stranded portion of a stem-loop region of the predicted secondary structure, the calculated free energy value for such stem-loop region is more positive than the free energy value calculated in step (b).

12. The method of Claim 11 wherein the free energy calculated in step (b) is in the range of from zero to about -7.0 kcal/mole.

13. The method of Claim 11 wherein the AUG initiation codon and the Shine-Dalgarno sequence are not included in a double stranded portion of a stem-loop

-34-

region of the predicted secondary structure of the modified mRNA.

14. The method of Claim 11 wherein the modification of the DNA or RNA sequence results in the production of at least five times more protein than produced by the unmodified DNA or RNA sequence.

15. A DNA or RNA sequence which has been modified by the method of Claim 11.

16. A transformed host cell which includes the modified DNA or RNA sequence of Claim 15.

FIG. I

FIG. 2

pNL014

pIN-1

XbaI +
partial EcoRI

XbaI + EcoRI

0.58 Kb fragment

Large fragment

Ligation

PNL015

coding region for
9aa

lppP   Met-α1   lppT

pNL015

ATG AAA GGG GGA ATT CCA AGC TTA TCG
MET LYS GLY GLY ILE PRO SER LEU SER

FIG. 3

pNL015

↓ Cla I

↓ S1

↓ EcoRI Linker+Ligase   pIN-1-A2

↓ EcoRI   ↓ EcoRI

Small Fragment

pNL008

ーlpp^P ———————————— Met-α1   lpp^T ーー

ATG AAA GGG AAG GAATTC
Met Lys Gly Lys Glu Phe

FIG. 4

FIG. 5

1   2
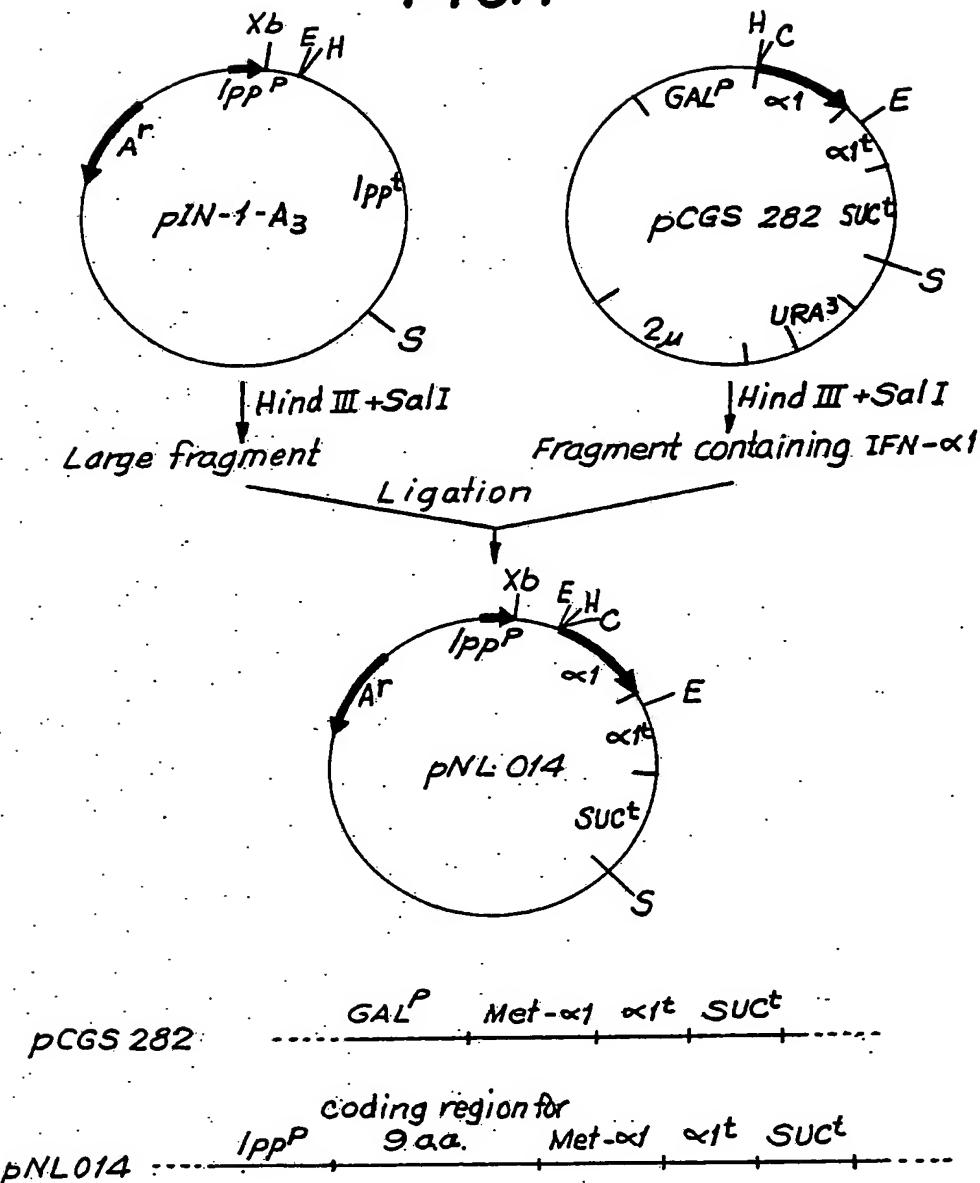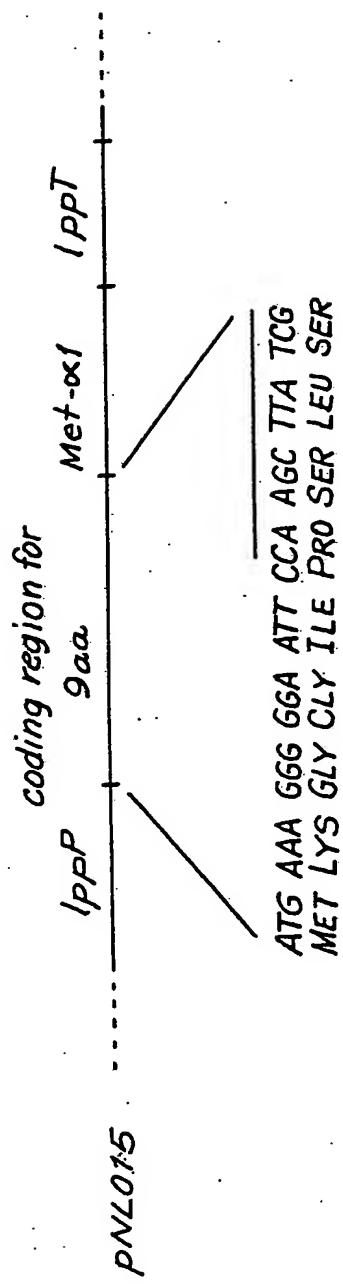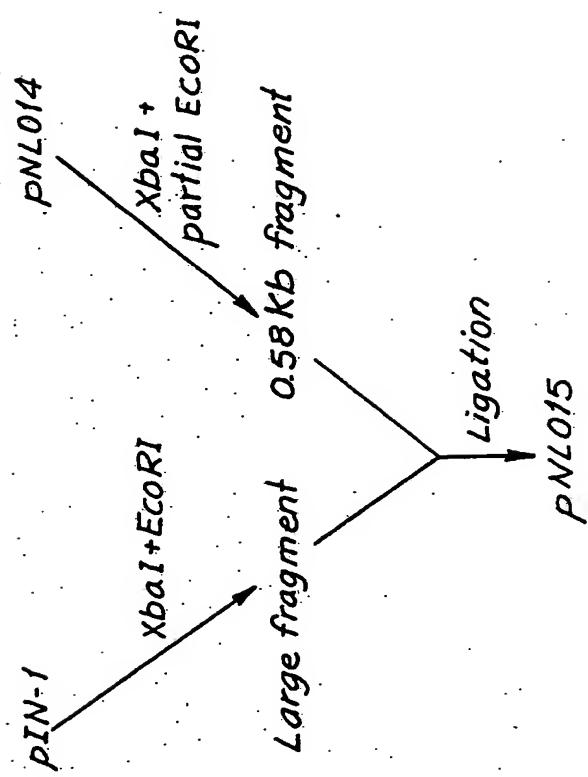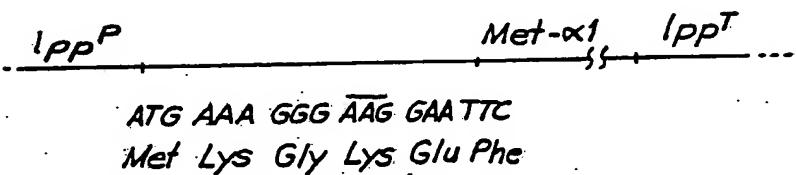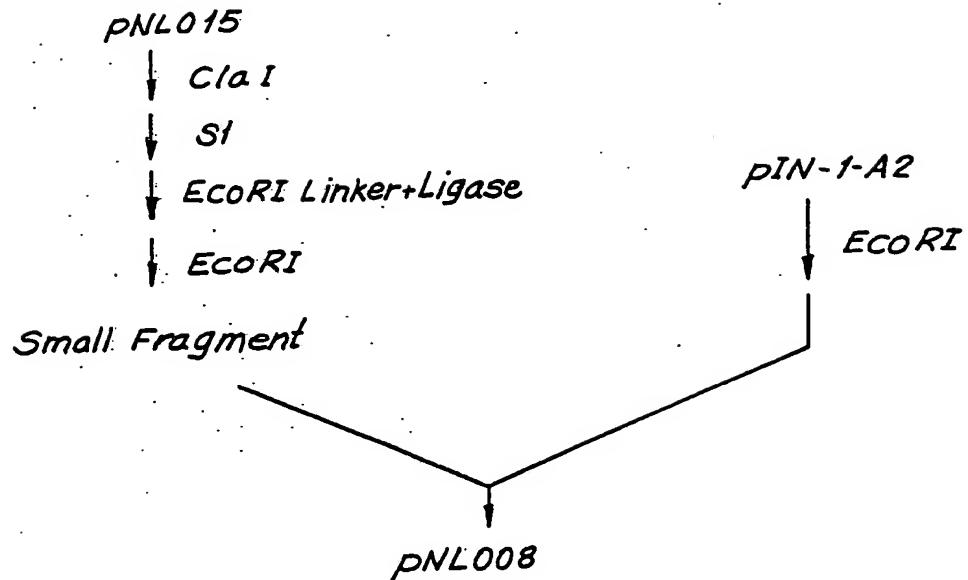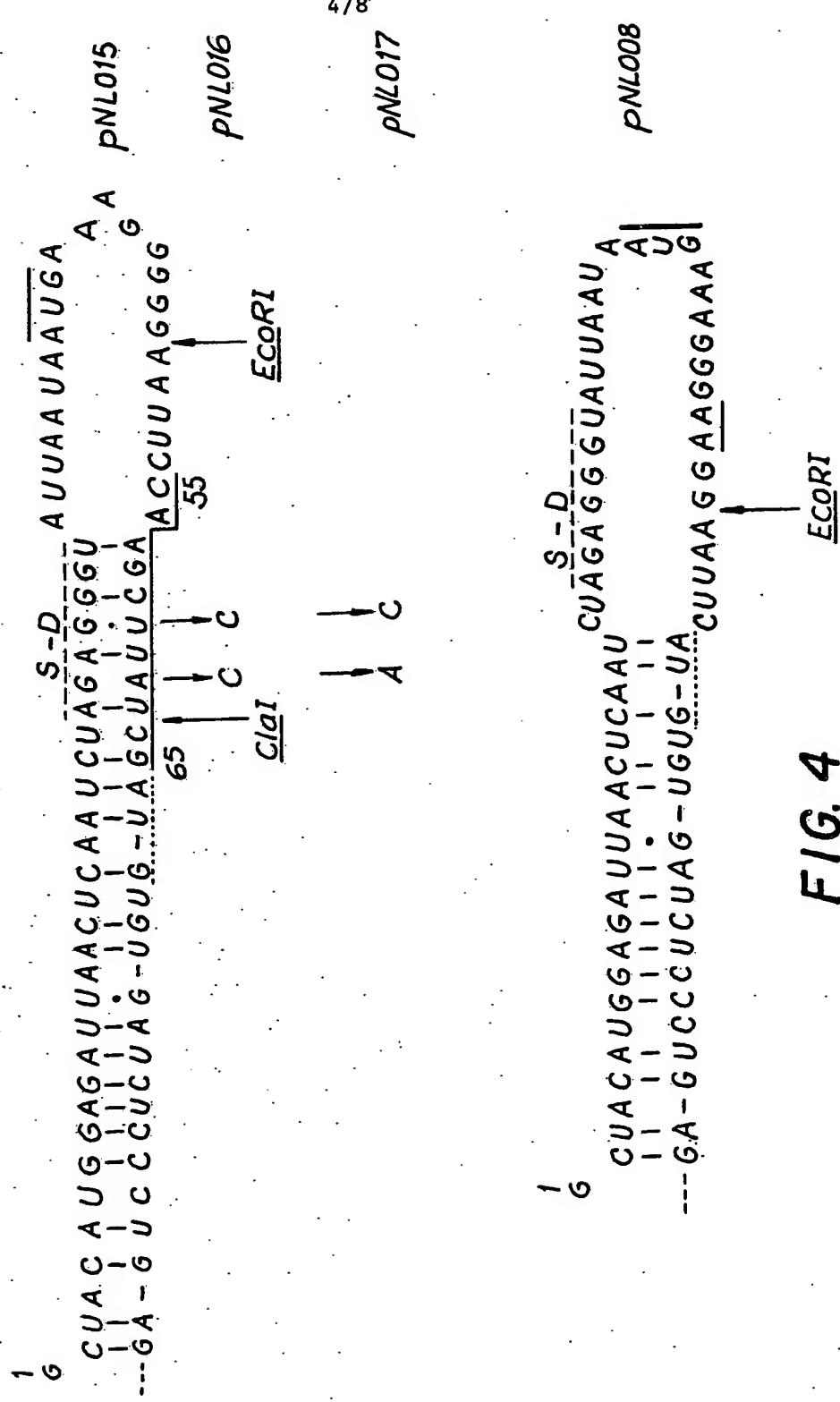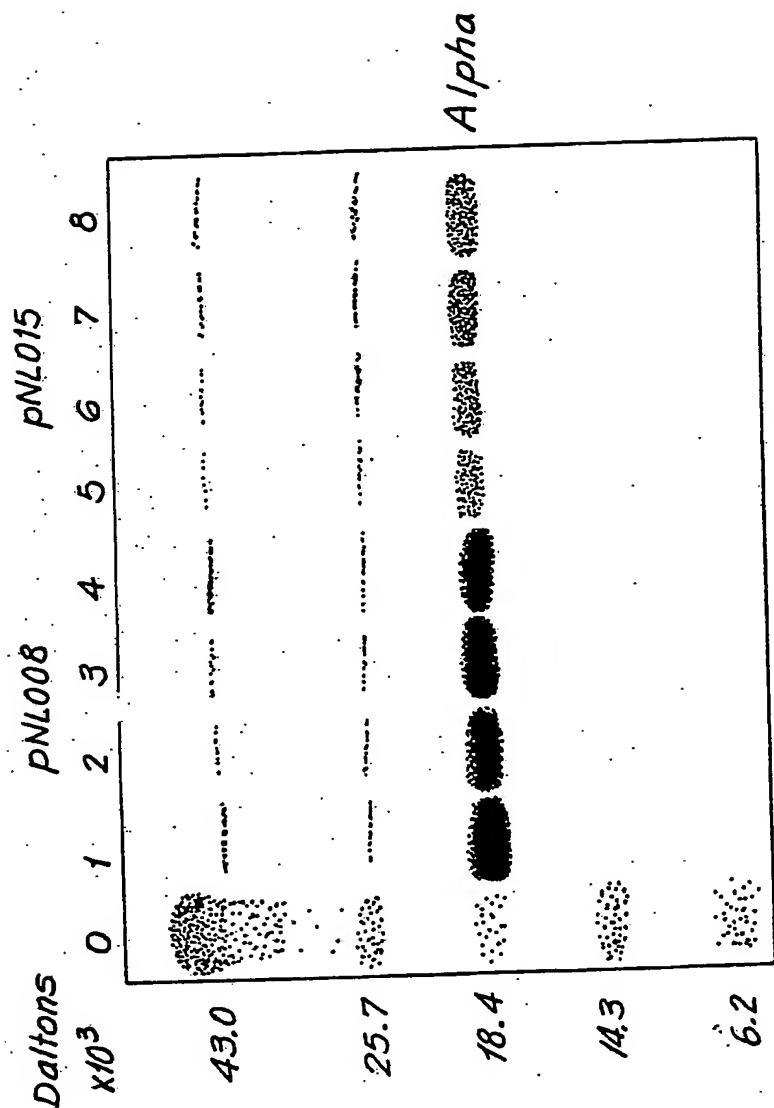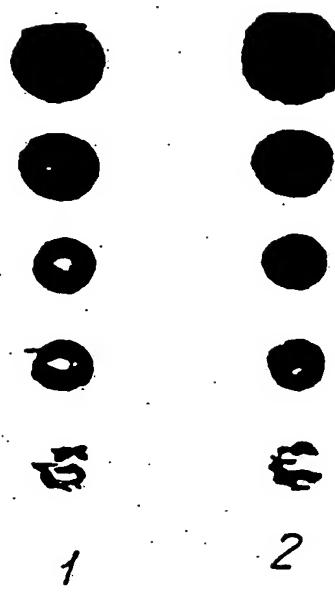
FIG. 6
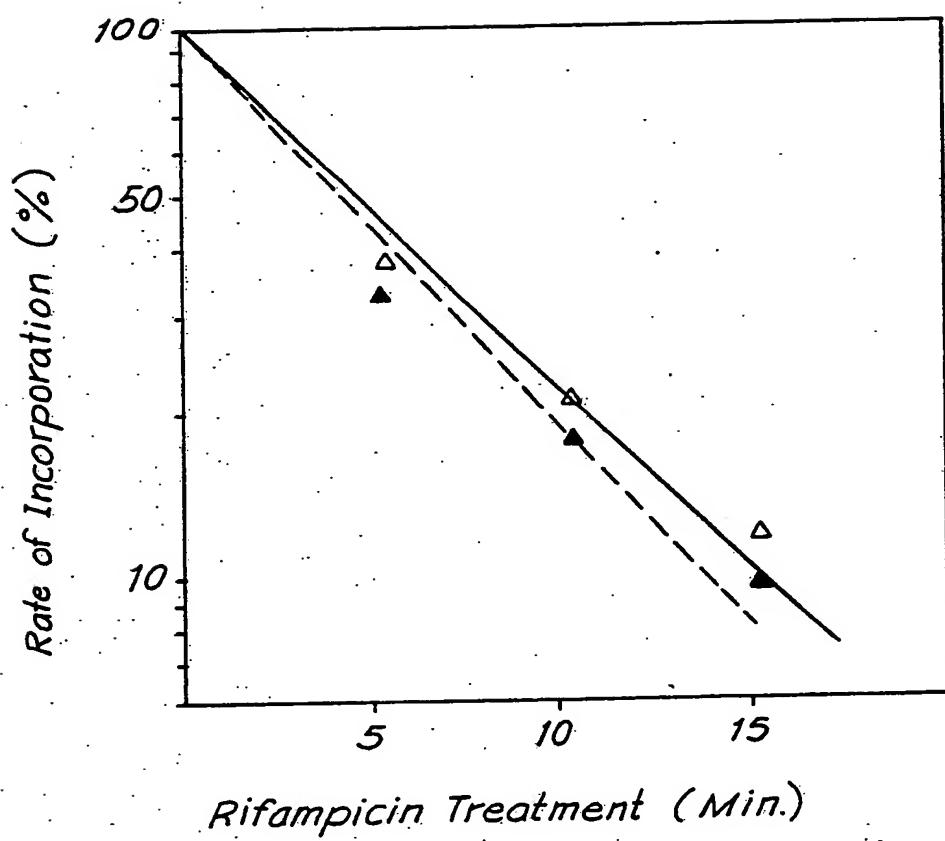
FIG. 7

FIG.8
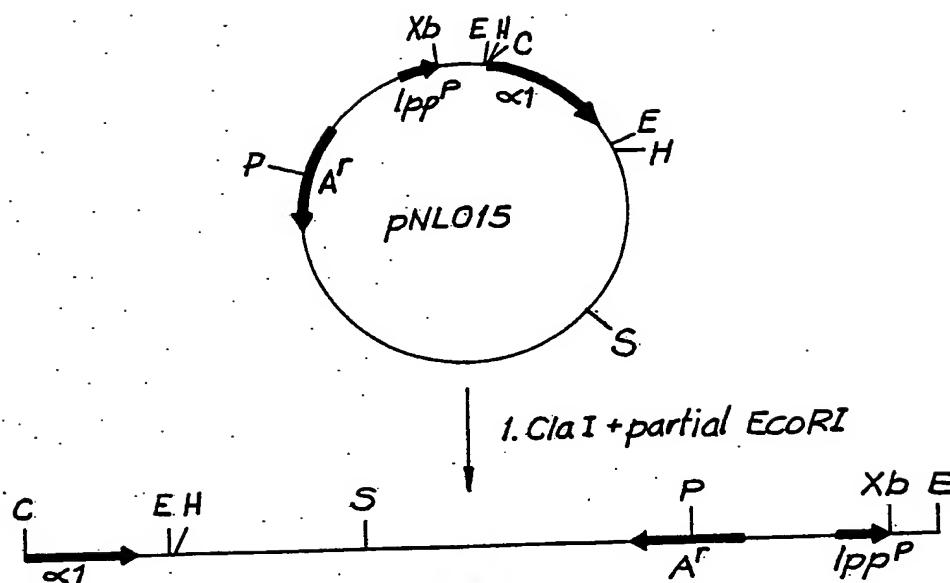
A: 5' AATTCCAAGCCTCT
3' GGTTCGGAGAGC

B: 5' AATTCCAAGCCTAT
3' GGTTCGGATAGC

1. Cla I + partial EcoRI

2. Add synthetic DNA fragment A or B + Ligase

3. EcoRI

4 Purify

5. Ligase

pNL016 or pNL017
(with fragment A)   (with fragment B)

INTERNATIONAL SEARCH REPORT

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) 6

According to International Patent Classification (IPC) or to both National Classification and IPC.

IPC(4): C12N 15/00; C07H 15/12; C12N 1/20
US CL.: 435/172.1; 435/172.3; 435/253; 536/27

II. FIELDS SEARCHED

Minimum Documentation Searched 7

Classification System	Classification Symbols
U.S.	435/68,172.1, 172.3 and 253 536/27 935/45,46

Documentation Searched other than Minimum Documentation
to the Extent that such Documents are Included in the Fields Searched 8

Chemical Abstracts Data Base (CAS) 1967-1988; Biological
Abstracts Data Base (BIOSIS) 1967-1988; Keywords:Stem Loop;
palindrome; inverted repeating sequences;mutation and shine dalgarno

III. DOCUMENTS CONSIDERED TO BE RELEVANT 9

Category *	Citation of Document, 11 with indication, where appropriate, of the relevant passages 12	Relevant to Claim No. 13
Y,P	US, A, 4,716,217 (CARUTHERS ET AL) 29 December 1987. See the entire document and in particular columns 10 and 11.	1-16
Y,P	GENE, Volume 58, issued July 1987 (Amsterdam, Netherlands), (LEE ET AL), "Modification of mRNA secondary structure and alteration of the expression of human interferon alpha 1 in Escherichia coli", see pages 77-86.	1-16
Y	JOURNAL OF BACTERIOLOGY, Volume 167, issued 2 September 1986, (Washington, D.C., USA), (AMBULOS ET AL), "Analysis of the regulatory sequences needed for induction of the chloramphenicol acetyltransferase gene cat-86 by chloramphenicol and amicetin", See pages 842-49.	1-16

* Special categories of cited documents: 10

"A" document defining the general state of the art which is not considered to be of particular relevance.

"E" earlier document but published on or after the International filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means.

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
11 October 1988	2 9 NOV 1988
International Searching Authority	Signature of Authorized Officer
ISA/US	Thomas D. Mays

Form PCT/ISA/210 (second sheet) (Rev.11-87)

III. DOCUMENTS CONSIDERED TO BE RELEVANT   (CONTINUED FROM THE SECOND SHEET)

Category *	Citation of Document. with Indication, where appropriate, of the relevant passages	Relevant to Claim No.
Y	NUCLEIC ACIDS RESEARCH, Volume 13, issued 25 March 1985 (Oxford, UK), (HALLEWELL ET AL), "Human Cu/Zn superoxide dismutase cDNA: Isolation of clones synthesising high levels of active or inactive enzyme from an expression library". see pages 2017-34.	1-16
Y	THE EMBO JOURNAL, Volume 4, issued September 1985, (Oxford, UK), (BRUCKNER ET AL), "Regulation of the inducible chloramphenicol acetyltransferase gene of the Staphylococcus aureus plasmid pUB112", see pages 2295-2300.	1-16
Y	MGG, Volume 182, issued July 1981 (BERLIN, FRG), (HORINOUCHI ET AL), "The control region for erythromycin resistance: Free energy changes related to induction and mutation to constitutive expression", see pages 341-348.	1-16